BRIEF TRAINING FOR PERFORMANCE OF A POINT ESTIMATION SONIFICATION TASK

Bruce N. Walker and Michael A. Nees

Sonification Lab
School of Psychology
Georgia Institute of Technology
654 Cherry Street
Atlanta, Georgia, USA 30332
bruce.walker@psych.gatech.edu, gtg673z@mail.gatech.edu

ABSTRACT

This study examined different types of brief training for a point estimation task with auditory graphs. Participants estimated the price of a stock at a specific times in a 10-hour trading day as depicted in a sonified graph of the stock price data. Forty Georgia Tech undergraduates completed a pre-test, an experimental training session, and a post-test for the point estimation task. In an extension of Smith and Walker [1], a highly conceptual, task analysis-derived method of training was compared to training paradigms that used either prompting of correct responses or feedback for correct answers during training. Two additional groups, one receiving only practice as training and another completing a filler task, were also included. Results indicate that practice with feedback for the point estimation task produced better post-test performance than all other training conditions.

1. INTRODUCTION

Visual depictions of information are pervasive (Jones and Careras [2] calculated that 2.2 trillion graphs were published in 1994 alone), and the average user has years of both formal and informal instruction with visual displays. Although sound was not a reasonable option for information display in most traditional analog systems, the advent of digital technology has made both the production and implementation of sounds feasible with any computer. As a result, auditory displays have been applied across numerous tasks in academic, military, medical, and industrial settings in recent decades (for a review, see [3]). Auditory displays show promise not only as alternatives or accompaniments to visual displays, but also in situations where vision is impaired, occupied, or obscured [1]. Non-speech audio may be a robust and effective means of encoding information in non-visual displays, but the novelty of audio representations may pose a significant obstacle to the widespread application of sound in information displays. The portrayal of quantitative information through sound is a relatively unfamiliar phenomenon, and novice users of auditory displays would benefit from an empirically established process to facilitate learning. Although researchers have recognized the importance of studying instruction for auditory displays (e.g., [3-5]) empirical studies of training for auditory displays remain largely overlooked.

The successful comprehension of a sound display requires perceptual acuity, and research has indicated that performance on perceptual tasks is susceptible to improvement with training (e.g., [6-12]). With regards to methods of improving perceptual skill such that the desired or correct response is associated with a perceptual stimulus, prompting of correct responses and feedback for correct responses have been compared extensively. Prompting involves training with a cue that provides the correct response paired to a perceptual stimulus just before or during the presentation of that stimulus. Feedback, on the other hand, requires the trainee to commit to a response, before the correct answer is revealed.

Sidley et al. [11] found no difference between prompting and feedback methods during training to identify complex auditory stimuli. Other research comparing prompting and feedback in perceptual learning resulted in mixed outcomes. Swets et al. [12] trained listeners to identify sounds that varied along five dimensions, and found that a group trained by prompting performed better than those trained by feedback. Aiken and Lau [7] reviewed a number of perceptual learning studies, and judged prompting to be at least as good as feedback across studies and better than feedback in many reports.

While perceptual learning literature should inform the design of a training regime for auditory displays, few studies have explicitly examined training for auditory displays. Lau [13] drilled sonar operators repeatedly to improve performance for a discrimination task with auditory sonar displays. Colquhon [14] used a signal detection sonar monitoring task to compare auditory, visual, and audiovisual sonar displays. Participants were permitted hours of practice prior to the testing sessions, but the author does not describe the details of the training. Similarly, Kobus et al. [15] showed that the best modality for accurate detection and categorization of sonar targets is dependent upon the properties of the target. Kobus et al. described their sample as "highly trained" (p. 24), but the nature of the extensive experience with the displays is not discussed.

The current study expands upon the findings of Smith and Walker [1] with regards to auditory graphs and training. Auditory graphs are a specific type of auditory display that only recently has begun to receive attention in empirical literature. In an auditory graph, quantitative data are assigned sounds, or sonified. The visual Y-axis value of a data point is typically translated to pitch in an auditory graph, and the data point's position along the visual X-axis is represented by the temporal presentation of the sound.

Research with auditory graphs has indicated that in many circumstances sound can provide as much information about the data in a graph as vision or touch. An early study of sonified graphs found them to be fairly comparable in efficacy to the tactile displays traditionally used to present quantitative information to the blind [5]. Flowers and Hauer [16] demonstrated that participants categorize groups of visual graphs and their auditory counterparts as perceptually similar or dissimilar along the same dimensions, namely shape, slope, and degree of linearity. Another study by Flowers [17] gave further evidence for the comparability of auditory and visual presentations of scatterplots. Participants estimated the same Pearson correlation r value for visual scatterplots and their corresponding auditory graphs. In general, people can both match an auditory graph to its visual counterpart [4] and produce a visual rendition of a graph that is over 80% accurate (on average) given a sonified presentation [18].

Smith and Walker [1] performed a pair of studies that looked at the effects of both auditory context and training on a point estimation task with auditory graphs. Of relevance to the current study, the researchers found that a group that received a training session performed significantly better on the task than did a control group that received no training.

Smith and Walker [1] used a point estimation task analysis as a conceptual basis for their training; their task analysis was consistent with existing models of visual graph comprehension (e.g., [19, 20]). Smith and Walker identified five steps essential to the successful completion of the point estimation task. Auditory graph users must: (a) listen to the data sonification, (b) identify the temporal position of the data point in question, (c) compare the data point to a baseline or reference tone, (d) estimate the magnitude of the difference (see [21]), and (e) assign a value to the data point based on arithmetic calculation. During training, Smith and Walker gave participants strategies for accomplishing each stage of the task, and practice trials with feedback were provided.

2. METHOD

Given that Smith and Walker [1] only looked at the effects of one type of training as compared to no training (a filler task), the current study took the next logical step and examined three qualitatively different types of training. In addition to Smith and Walker's conceptual, task analysis-derived method of training, the current study examined two perceptual training techniques (prompting and feedback). Another training condition included only practice (without prompting, feedback, or conceptual training) with the task. Finally, a control condition involved a filler task during training.

2.1. Participants

Forty sighted Georgia Tech undergraduates participated. The sample consisted of 13 males and 27 females (mean age = 20.27 years, *SD* = 1.68). All participants reported normal or corrected-to-normal hearing and vision.

2.2. Apparatus

All visual presentations including instructions and visual components of training (where applicable) were made on a 17 in. (43.2 cm) Apple Studio Display monitor. All auditory presentations were delivered via Sony MDR-7506 headphones adjusted for fit and comfortable volume. Data for auditory graphs were sonified as MIDI files using the Sonification Sandbox program [22, 23], converted to .WAV files, and imported to SoundForge XP for track mixing. All sounds presented the same signal to both left and right headphone channels. The experiment was programmed and conducted in Macromedia Director version 8.5.

2.3. Stimuli

Stimuli consisted of auditory graphs depicting the price of a stock in dollars as it fluctuated over the course of a 10-hour trading day, opening at 8 a.m. and closing at 6 p.m. The price of the stock in dollars (on the Y-axis) was represented by discrete tones that changed in pitch as the price changed, while each hour of the trading day (X-axis) corresponded to one second in time. Discrete tones were presented at the rate of two per second, whereby the pitch of each tone corresponded to the price of the stock at each half-hour of the trading day.

The frequencies of the tones were matched to the price of the stock in dollars using the Sonification Sandbox [22, 23], and frequencies used ranged from 196 Hz for the lowest sonified stock price dollar value to 1979.5 Hz for the highest data value. Data values in between the maximum and minimum were sonified on an exact scale, whereby data values falling between notes were adjusted in pitch (rather than rounded to the nearest musical note) by the software to represent the exact frequency of the data point on the scale. Stock data were sonified using the piano from the MIDI instrument bank.

Each auditory graph was given Y-axis (price) auditory context using a dynamic Y reference tone that brackets sets of consecutive ascending or descending tones. The Y reference is a beeping tone displayed concurrently with the actual discrete data points of the auditory graph. The dynamic Y reference tone creates the auditory equivalent of visual gridlines for Y-axis values. The tone represented the stock's maximum price of the day ($84) when the true price of the stock was ascending and indicated the stock's lowest price of the day ($10) when the price was descending. In the current study, maximum value Y context was sonified using a piccolo timbre, while minimum value Y context was created with a bassoon timbre. X-axis (time) context was provided by the addition of a track that featured a rhythmic acoustic snare beat every 1 s of the display (i.e., on every hour). The dynamic Y reference has been shown to generally provide helpful context for point estimation tasks with auditory graphs, and the X-axis click track is helpful when the density of the track is different than the data density and thus provides added information [1, 24].

During the pre-test phase of the study, the auditory graph for all conditions consisted of a pseudo-sinusoidal graph whose values rose for the first three hours of the trading day, fell for the next five hours, and rose for the last three hours. The stock opened and closed at a price of 50 dollars. As in Smith and Walker [1], the pre-test graph was presented devoid of all context such that participant heard only the sonified data points. During the post-test phase, participants in all conditions again heard the same auditory graph, but the post-test graph was enriched with both X-axis context and Y-axis context.

2.4. Procedure and Task

Participants were randomly assigned to one of five experimental training conditions. Demographic information was collected before the pre-test session of the study. Participants received instructions that included a brief description of the auditory graph and task. The pre-test was the same for all participants, and the pre-test auditory graph did not employ context. In order to provide a baseline reference, participants were told that the opening price of the stock was 50 dollars. For the 11 pre-test trials, participants were asked to identify the price of the stock for each hour (8 a.m. – 6 p.m.) of the trading day in a randomly selected order. A single trial began with a visual text presentation of the test question (e.g., "What is the price of the stock at 10 a.m.?") followed by the presentation of the auditory graph. Participants were allowed to listen to the auditory graph as many times as needed before responding. Participants were given a two-minute break following the pre-test.

After the break, participants completed the experimental training manipulation for their assigned conditions as described below. The training sessions were designed to be self-paced and approximately equivalent in average duration for each of the five groups. The training portion of the study lasted between 20 and 25 minutes, and participants received a two-minute break following completion of training. A different auditory graph was used during the training phase of the study (with the exception of Group 1, a control group that experienced no sonifications during training) to prevent the simple memorization of stock price values from confounding the post-test data, particularly in those experimental conditions where the exact values of the stock price were provided to participants during training via feedback or prompting.

The post-test segment of the study began following the final break. All participants were given the same task from the pre-test, where the price of the stock was estimated for each hour of the trading day over 11 trials of randomly selected hours. The post-test instructions also included a brief description of the post-test auditory graph, which did employ context and was the same for all conditions. As in Smith and Walker [1], the dependent variable was defined as the root mean squared (RMS) error (in dollars) of participants' responses to the point estimation trials.

2.5. Experimental Training Conditions

Group 1 was a control group that experienced a filler task during the training manipulation. Participants heard a recorded passage about tiger sharks and were required to respond to questions related to the passage. By using a spoken-word passage, the auditory nature of the filler task maintained consistency with other training conditions.

Group 2 experienced only practice with auditory graphs as training. Participants underwent 22 trials (two complete sets of each hour in the trading day) of the point estimation task during the training session without feedback, prompting of correct responses, or conceptual training.

Group 3 had 22 trials of the point estimation task with the aid of feedback regarding performance during the training session. Participants listened to the training auditory graph and made a response; disclosure of the correct response followed.

Group 4 had 22 trials of the point estimation task with the guidance of a visual prompt during training. Participants heard the training graph, and the visual display concurrently presented text indicating the time of day and stock price.

Group 5 experienced the same training procedures as described in Smith and Walker [1]. An interactive presentation with a voice-over and visual aids explained sonification and auditory graphs. Participants were given detailed information about the nature of the auditory graph and its contextual elements, which were equated with familiar contextual cues in visual graphs. The point-estimation task was broken down into its component steps during training, and participants were given strategies for accomplishing each step of the task successfully. Participants practiced determining the part of the sonification representing a given time of day, and a second set of practice trials emphasized estimating the magnitude of the stock price.

3. RESULTS

Participants' post-test RMS error scores were analyzed with a one-way, univariate analysis of covariance (ANCOVA). Participant pre-test scores were entered into the analysis as the covariate, and the type of training was examined as a between subjects factor. The ANCOVA revealed significant between groups effects, F (4, 34) = 3.22, p <.05. Adjusted mean RMS error scores for the post-test are presented in Table 1.

Table 1. *Adjusted mean post-test RMS scores.*

Group	Training session manipulation	Adjusted RMS error ($)	S.D. ($)
1	Filler task	25.04	10.6
2	Practice only	22.47	10.2
3	Practice with feedback	8.45	10.2
4	Practice with visual prompt	14.38	10.1
5	Conceptual training	18.00	10.1

Tukey's HSD test revealed a significant difference between Groups 1 and 3, p < .05. Note that there are more data currently being gathered, and we anticipate that the groups will be more clearly differentiated with those data included.

4. DISCUSSION

The current study examined the effects of qualitatively different types of training for a point estimation task with auditory graphs. Although Group 5 was trained to successfully accomplish each of the individual stages of the point estimation task, Group 3 was the only condition to exhibit significantly superior post-test performance over the control group. Group 3 was trained with feedback, a technique that has been employed in previous perceptual learning research. Intuitively, this pattern of results indicates that the perceptual elements of the task, namely judging the values corresponding to pitches of tones, may be more difficult to accomplish than other aspects of the task and deserve special attention in an ideal training session. Even in the initial stage of exposure to auditory graphs, the benefit of providing participants with a conceptual tutorial for auditory graphs was overshadowed by the naïve users' need to simply practice the task with feedback regarding the correct responses.

These results also suggested a trend whereby training conditions featuring either a visual prompt or conceptual training were approximately equal in efficacy and better than no training or practice alone, but statistical significance was not obtained. We expect that any treatment effects will become clearer and more pronounced when data from an expanded sample of 60 participants have been obtained and analyzed.

The successful deployment of auditory graphs will require effective and efficient training programs. While conceptual training has been shown to be effective, other approaches such as practice with feedback may be even more effective. The training session in the current study was a single, brief session, and accuracy (as measured by RMS error) was the only outcome measured. Further research should examine training over multiple sessions in order to determine an optimal training regime for speed and accuracy. In all likelihood, the type of training that best facilitates improved performance will change across stages of skill acquisition. Ultimately, researchers will need to examine transfer of training across tasks (point comparison, trend evaluation, extrapolation, etc.) with auditory graphs, and a successful training regime for auditory graphs should be adapted for other types of auditory displays.

5. REFERENCES

[1] D. R. Smith and B. N. Walker, "Effects of auditory context cues and training on performance of a point estimation sonification task," *Journal of Applied Cognitive Psychology*, in press.

[2] R. W. Jones and I. E. Careras, "The empirical investigation of factors affecting graphical visualizations.," *Behavior Research Methods, Instruments & Computers*, vol. 28(2), pp. 265-269, 1996.

[3] G. Kramer, B. N. Walker, T. Bonebright, P. Cook, J. Flowers, N. Miner, J. Neuhoff, R. Bargar, S. Barrass, J. Berger, G. Evreinov, W. T. Fitch, M. Gröhn, S. Handel, H. Kaper, H. Levkowitz, S. Lodha, B. Shinn-Cunningham, M. Simoni, and S. Tipei, "The Sonification Report: Status of the Field and Research Agenda. Report prepared for the National Science Foundation by members of the International Community for Auditory Display," International Community for Auditory Display (ICAD), Santa Fe, NM 1999.

[4] T. L. Bonebright, M. A. Nees, T. T. Connerley, and G. R. McCain, "Testing the effectiveness of sonified graphs for education: A programmatic research project.," presented at The Seventh International Conference on Auditory Display, Espoo, Finland, 2001.

[5] D. L. Mansur, M. M. Blattner, and K. I. Joy, "Sound graphs: A numerical data analysis method for the blind.," *Journal of Medical Systems*, vol. 9, pp. 163-174, 1985.

[6] M. Ahissar, "Perceptual training: A tool for both modifying the brain and exploring it.," *Proceeding of the National Academy of Sciences*, vol. 98, pp. 11842-11843, 2001.

[7] E. G. Aiken and A. W. Lau, "Response prompting and response confirmation: A review of recent literature.," *Psychological Bulletin*, vol. 68, pp. 330-341, 1967.

[8] J. Annett, "Training for perceptual skills," *Ergonomics*, vol. 9, pp. 459-468, 1966.

[9] J. Annett and L. Patterson, "Training for auditory detection.," *Acta Psychologica*, vol. 27, pp. 420-426, 1967.

[10] D. C. Prather, G. A. Berry, and J. M. Bermudez, "The effect of prompting and feedback on performance during learning, stress, and transfer of a perceptual skill.," *Proceedings of the Annual Convention of the American Psychological Association*, vol. 7, pp. 643-644, 1972.

[11] N. A. Sidley, E. Winograd, and E. W. Bedarf, "Learning to identify complex sounds: Prompting versus confirmation," *Journal of the Acoustical Society of America*, vol. 38, pp. 1050-1052, 1965.

[12] J. A. Swets, S. H. Millman, W. E. Fletcher, and D. M. Green, "Learning to identify nonverbal sounds: An application of a computer as a teaching machine.," *Journal of the Acoustical Society of America*, vol. 37(7), pp. 928-935, 1962.

[13] A. W. Lau, "Descriptive analysis of Doppler discrimination as a function of variations in dimensions of the sonar echo.," *Journal of the Acoustical Society of America*, vol. 40, pp. 565-569, 1966.

[14] W. P. Colquhon, "Evaluation of auditory, visual, and dual-mode displays for prolonged sonar monitoring in repeated sessions," *Human Factors*, vol. 17, pp. 425-437, 1975.

[15] D. A. Kobus, L. M. Ward, J. R. Hawker, N. A. Sidley, E. Winograd, E. W. Bedarf, T. Karlsson, P. N. Chase, S. Kalyuga, P. Chandler, J. Sweller, M. Ahissar, S. Hochstein, P. David, E. Hirshman, S. Tindall-Ford, M. Dubois, I. Vial, J. Meyer, E. G. Aiken, and A. W. Lau, "Multimodal detection and recognition performance of sonar operators," *Human Factors*, vol. 28, pp. 23-29, 1986.

[16] J. H. Flowers and T. A. Hauer, "Musical versus visual graphs: Cross-modal equivalence in perception of time series data.," *Human Factors*, vol. 37, pp. 553-569, 1995.

[17] J. H. Flowers, D. C. Buhman, and K. D. Turnage, "Cross-modal equivalence of visual and auditory scatterplots for exploring bivariate data samples.," *Human Factors*, vol. 39, pp. 341-351, 1997.

[18] L. M. Brown and S. A. Brewster, "Drawing by Ear: Interpreting Sonified Line Graphs," presented at International Conference on Auditory Display ICAD2003, Boston, MA., 2003.

[19] P. A. Carpenter and P. Shah, "A model of the perceptual and conceptual processes in graph comprehension," *Journal of Experimental Psychology: Applied*, vol. 4, pp. 75-100, 1998.

[20] D. J. Gillan and R. Lewis, "A componential model of human interaction with graphs: I. Linear regression modeling.," *Human Factors*, vol. 36, pp. 419-440, 1994.

[21] B. N. Walker, "Magnitude estimation of conceptual data dimensions for use in sonification.," in *Journal of Experimental Psychology: Applied*, vol. 8: American Psychological Assn, 2002, pp. 211-221.

[22] B. N. Walker and J. T. Cothran, "Sonification Sandbox: A graphical toolkit for auditory graphs," presented at Ninth International Conference on Auditory Display ICAD2003, Boston, MA, 2003.

[23] B. N. Walker and M. Lowey, "Sonification Sandbox: A graphical toolkit for auditory graphs," presented at Rehabilitation Engineering & Assistive Technology Society of America (RESNA) 27th International Conference, Orlando, FL, 2004.

[24] D. R. Smith and B. N. Walker, "Tick-marks, axes, and labels: The effects of adding context to auditory graphs," presented at Eighth International Conference on Auditory Display, Kyoto, Japan, 2002.